Belgian Medical Genomics Initiative

BeMGI 2014 Annual Meeting : Poster presentations

Large-scale analysis of tandem repeat variability in the human genome

Alena Zablotskaya[1,2,3], Jorge Duitama[4,5], Rita Gemayel[4], An Jansen[2,3,4], Stefanie Belet[2,3], Joris R. Vermeesch[1], Kevin J. Verstrepen[4] and Guy Froyen[2,3]

[1]Laboratory for Cytogenetics and Genome Research, Center for Human Genetics, University Hospitals Leuven, KU Leuven, B-3000 Leuven, Belgium, [2]Human Genome Laboratory, VIB Center for the Biology of Disease, Leuven, Belgium, [3]Human Genome Laboratory, Department of Human Genetics, KU Leuven, B-3000 Leuven, Belgium, [4]VIB lab for Systems Biology & CMPG Lab for Genetics and Genomics, KU Leuven, B-3001 Leuven, Belgium, [5]Agrobiodiversity Research area, International Center for Tropical Agriculture (CIAT), Cali, Colombia

ABSTRACT

Tandem repeats are short DNA sequences that are repeated head-to-tail with a propensity to be variable. They constitute a significant proportion of the human genome, also occurring within coding and regulatory regions. Variation in these repeats can alter the function and/or expression of genes allowing organisms to swiftly adapt to novel environments. Importantly, some repeat expansions have also been linked to certain neurodegenerative diseases. Therefore, accurate sequencing of tandem repeats could contribute to our understanding of common phenotypic variability and might uncover missing genetic factors in idiopathic clinical conditions. However, despite long standing evidence for the functional role of repeats, they are largely ignored because of technical limitations in sequencing, mapping and typing. Here, we report on a novel capture technique and data filtering protocol that allowed simultaneous sequencing of thousands of tandem repeats in the human genomes of a three generation family using GS-FLX-plus Titanium technology. Our results demonstrated that up to 7.6% of tandem repeats in this family (4% in coding sequences) differ from the reference sequence, and identified a de novo variation in the family tree. The method opens new routes to look at this underappreciated type of genetic variability, including the identification of novel disease-related repeats.